

Ulohy k predmetu Pocitacova statistika

Jan Somorcik
FMFI UK Bratislava

22. februára 2016

1 Testy o strednej hodnote

1. Dedo Mraz a vodka

Dedo Mraz roznasa kazdy rok moskovskym detom darceky. Niektore roky pocas roznasania popija vodku. V subore `dedo_mraz.txt` su zaznamenane udaje za poslednych 200 rokov: stlpec `cas` hovorí, kolko hodin roznasanie trvalo a stlpec `popijal` hovorí, ci Dedo Mraz popijal. Dedo Mraz rokmi nadobudol tusenie, ze ak popija vodku, roznasanie mu ide rychlejsie.

Ilustrujte Dedove tusenie boxplotom a otestujte ho.

2. Liek zahustuje krv

Farmaceuticka firma vyvinula novy liek proti depresii. Objavilo sa vsak podozrenie, ze jeho neziaducim ucinkom je kratkodobe zahustenie krvi. Preto sa liek otestoval na 29 ludoch (subor `zahustenie_krvi.txt`): kazdemu sa zmerala hustota krvi pred (stlpec `pred`) a po (stlpec `po`) uziti lieku.

Zahustenie krvi ilustrujte obrazkom a otestujte, ci je statisticky vyznamne.

3. Formula 1

Stajna Williams pouzivala doteraz pneumatiky Bridgestone, no nebola spokojna. Preto si zadovazili zopar pneumatik Michelin a testovaci pilot Alexander Wurz s nimi jazdil testovacie jazdy na okruhu Silverstone. V tabulke `formula1.txt` su zapisane dosiahnute casy na 1 kolo (stlpec `casy`) a ake pneumatiky sa pouzili (stlpec `pneumatiky`).

Ktore pneumatiky sa zdaju byt lepsie? Tusenie dolozte obrazkom a otestujte. Co vysledok testu prakticky znamena pre stajnu Williams?

4. Smrek

Lesnici nahodne vyrubali zopar smrekov v narodnom parku TANAP a v narodnom parku Slovensky Raj. V subore `smrek.txt` su zaznamenane vysky tychto smrekov a kde boli vyrubane.

Kde sa zda, ze rastu vyssie smrek? Tusenie dolozte obrazkom a otestujte. Vysledok ozrejmite „ludskou recou“.

5. Sprinteri a doping

Americke laboratorium BALCO lekara Victora Conteho vyvinulo doping THG pre spinterov. Po prepuknutí skandalu nasla FBI subor `sprint.txt`, kde boli uvedene tajne kody sprinterov (=code) a casy v behu na 100m pred (=before) a po (=after) uziti THG.

- a) U kolkych sprinterov malo THG dobry a u kolkych zly vplyv na vykon?
- b) Zistite kod sprintera s najlepsim casom pred dopovanim.
- c) Zistite kod sprintera, u ktoreho doslo k najvacsiemu zlepšeniu vykonnosti.
- d) Otestujte tusenie, ze THG zlepšuje vykonnosť. „Ludskou recou“ skomentujte vysledok.

6. Prepitne a euro

Casnicka skumala, ako sa zmenila vyska prepitneho po zavedeni eura. Pocas dvoch podobnych dni si zapisovala vysku prepitnych: v subore `prepitne.txt` su v stlpci `tringelt` uvedene vysky prepitnych a v stlpci `mane.euro` je poznamene, ci islo o den pred alebo po zavedeni eura. Prepitne z „neeuroveho“ dna su uvedene v korunach, z „euroveho“ v eurach, konverzny kurz je 1€=30,1260 SKK. Ovplynulo zavedenie eura vysku prepitneho? Svoje tusenie ilustrujte obrazkom, otestujte ho a vysledok ozrejmite „ludskou recou“.

7. „Zivotnost“ trenerov (len pre panov)

V subore `treneri.txt` su zozbierane udaje o poslednych 53 trenerskych vymenach v kluboch najvyssjej ceskej a slovenskej futbalovej ligy. Stlpec `doba` hovorí, kolko dni vydržal dany trener, kým ho odvolali alebo odišiel sam, stlpec `krajina` hovorí, či slo o cesky alebo slovensky klub. V ktorej krajine vydržia treneri dlhsie? Svoje tusenie podložte obrazkom, otestujte a vysledok vysvetlite „ludskou recou“.

8. Christian Dior (len pre damy)

Siet parfumerii FAnn vyhodnocovala predajnosť výrobkov značky Christian Dior vo svojich predajniach v Bratislave a v Prahe. V subore `christian_dior.txt` su zozbierane udaje zo zakaznickych kariet. Stlpec `suma` hovorí, kolko držiteľka karty minula (suma je v eurach) v uplynulom roku na výrobky spominanej značky. Stlpec `mesto` hovorí, či slo o Prazacku alebo Bratislavcanku. V ktorom meste su fanyňky Christian Diora ochotnejšie minat peniaze na svoju oblubenú značku? Svoje tusenie podložte obrazkom, otestujte a vysledok vysvetlite „ludskou recou“.

9. Znamky gymnazistov (autor: Jan Kudlacak)

V subore `znamky.txt` su koncorocne znamky z matematiky studentov Gymnazia sv. Cyrila a Metoda v Nitre (stlpec `znamka`). V stlpci `typ` je uvedene, ci je ziak na 8-rocnom alebo 4-rocnom studiu. Ktory typ ziaikov sa zda byt v matematike sikovnejši? Svoje tusenie ilustrujte obrazkom, otestujte a vysledok vysvetlite „ludskou recou“.

10. Piatok trinasteho

V subore `piatok_trinasteho.txt` su udaje o pocte osetrenych urazov v 20 nemocniciach na Slovensku v piatok 6. juna 2008 (stlpec `piatok6`) a v piatok 13. juna 2008 (stlpec `piatok13`).

- a) Zistite, v ktorej nemocnici prislo 13. juna 2008 k najvyssiemu percentuálnemu narastu poctu urazov oproti 6. junu 2008.
- b) Zda sa byt „piatok trinasteho“ na zaklade nasich udajov nestastnym dnom? Svoje tusenie otestujte pomocou p-hodnoty i pomocou kritických hodnot a vysledok vysvetlite „ludskou recou“.

11. Policajne radary

Policajti dostali novy radar na meranie rychlosti. Po par dnoch pouzivania nadobudli tusenie, ze pomocou noveho radaru lepsie zarabajú (t.j. novy radar hlasi v priemere vyssie hodnoty nez stary). Preto pri botanickej zahrade v Bratislave urobili experiment, pri ktorom merali rychlost oboma radarmi: v subore `policajne_radary.txt` je v kazdom riadku rychlost nejakeho auta namerana starym radarom i novym radarom.

- a) Vysledky merani interpretujte pomocou vhodnych obrazkov.
- b) Na zaklade dat otestujte tusenie policajtov a vysledok objasnite „ludskou recou“.
- c) Z rozdelenia s akym sklonom zrejme pochadzaju udaje v stlpci `novy`? Uvedte aspon 2 dovody.

12. Kuchynske vahy

Redakcia casopisu „Test pro spotrebitele“ porovnavala dve digitalne kuchynske vahy: ETA a Gorenje. Na kazdu z nich niekolkokrat polozili to iste zavazie a namerane hmotnosti zapisali do tabulky. (Ako je zname, namerana hmotnosť je ovplyvnená tým, *kam* presne sa polozi zataz.) Vysledky su uvedene v tabulke `kuchynske_vahy.txt`. Mozno vahy povazovat za odlisne z pohladu rozptylenia nameranych hodnot, resp. z pohladu priemernych vysledkov? Situáciu znazornite graficky, urobte testy a vysledky objasnite „ludskou recou“.

13. Propagacne zajazdy

Agentura ***** organizuje zajazdy lahko ovplyvritelnych starsich ludi: do otlcenenho autobusu rano zdarma nalozi 45 dochodcov a odvezie ich „akoze“ na vylet do nejakej nie prilis dalekej obce. Tam sa vylet nahle zmeni na nekonecnu predvazdaciú akciu velmi „potrebných“ doplnkov do domacnosti, ktore si je mozne hned aj kupit, ci objednat. V tabulke `propagacne_zajazdy.txt` su udaje z niekoľkych takychto seans. V stlpci `trzba` je uvedene, kolko € v dany den agentura utrzila, a v stlpci `rezen` je informacia, ci sa pocas danej akcie okrem bezneho obcerstvenia v podobe vlazneho caju a oschnutej salamy podavalo i obcas avizovane „A caka Vas gurmanske prekvapenie!“ v podobe miniaturneho reza. Otestujte, ci podavanie reza ovplyvnuje trzbu a vysledok ozrejmite „ludskou recou“.

BONUS: Oplati sa agenture podavat rezne, ak ich objedna po 2€ za porciu?

14. Billa card

Spolocnost Billa vyhodnotila predvianocne nakupy 100 najcastejsich pouzivatelov vernostnej karty Billa v Bratislave. V subore `billa_card.txt` su v stlpcoch `vianoce2008` resp. `vianoce2009` celkove hodnoty nakupov (prepoctane na €) tychto ludi pocas 7 dni pred Vianocami 2008 resp. 2009.

- Z dalsich analyz vyhodte zakaznikov, ktorí su outliermi z pohladu zmeny mnozstva minutych €.
- Otestujte, ci je mozne vyhlasiť, ze zakaznici sa pred Vianocami 2009 z dovodu krizy pri nakupoch potravín uskromnili. Vysledok ozrejmite „ludskou recou“.

15. Zmrzlina

Podnikatel cez letne prazdniny zamestnal vo svojich dvoch zmrlinovych stanokoch na namesti brigadnika a brigadnicku: v subore `zmrzlina.txt` predstavuje kazdy riadok jeden den prazdnin a uvedene cisla su denne trzby brigadnika (stlpec `muz`) a brigadnicku (stlpec `zena`).

- Tabulku trzieb interpretujte pomocou vhodnych obrazkov.
- Otestujte podnikatelovo tusenie, ze zakaznikov lepsie pritahuje brigadnicka a vysledok objasnite „ludskou recou“.
- Z rozdelenia s akym sklonom zrejme pochadzaju brigadnickine trzby? Uvedte aspon 2 dovody.

16. Strelba revolverom

Policajt-pravak cvicne strielal revolverom na terc. V subore

`strelba_revolverom.txt` su zaznamenané vodorovné odchylky zásahu od stredu terca (stĺpec `odchylka`) a to, v ktorej ruke policajt držal revolver (stĺpec `ruka`).

- a) Z každej zo sad vychylok (teda pre ľavú i pre pravú ruku) vypíšte outlierov a ďalšie analýzy robte bez nich .
- b) Je možné považovať strelbu pravou rukou za systematicky vychylenú?
- c) Porovnajte presnosti strelby pravou a ľavou rukou z pohľadu rozptylnosti a z pohľadu priemernej vychylenosti.

17. Hudba v Bille

Billa skúmala opodstatnenosť púšťania hudby v predajniach. V súbore `hudba_v_bille.txt` sú v každom riadku údaje z nejakej predajne, pričom `ticho` znamená tržbu v piatok, ak nebola zapnutá hudba, a `hudba` znamená tržbu v iný piatok, keď hudba zapnutá bola.

- a) Z ďalších úvah vyhodte predajne, v ktorých možno považovať zmenu tržby spôsobenú hudbou za outliera.
- b) Otestujte zmysluplnosť púšťania hudby v predajniach a výsledok ozrejmite „ľudskou recou“.

18. Sackové vs. flaskové mlieko

V TESCO predávajú okrem iných druhov polotučných mliek aj značku `Levmilk`, a sice v sákoch alebo plastových fľašiach. Keďže TESCO má od výrobcu z oboch balení rovnakú maržu, podstatný je len počet predaných kusov. V súbore `mlieko.txt` sú údaje z 30 dní, kedy ponúkali z druhu `Levmilk` vždy buď len sákové alebo len flaskové balenie (stĺpec `druh`). Počty predaných kusov sú v stĺpci `predaj`.

- a) Pre každý druh mlieka najprv vyhodte dni, v ktorých bol počet predaných mliek outlierom. Ďalšie výpočty robte bez týchto dní.
- b) Ktoré balenie sa míňa lepšie? Svoje tušenie otestujte a výsledok ozrejmite „ľudskou recou“.

19. SadloVrah

O prostriedku na chudnutie menom `SadloVrah`, ktorý vraj zaručí schudnutie bez zmeny životosprávy, koluje fáma: vraj ľudia od neho v skutočnosti priberajú! V súbore `SadloVrah.txt` sú údaje zozbierané od 62 dobrovoľníkov: stĺpec `pred` znamená hmotnosť pred užívaním a stĺpec `po` znamená hmotnosť po mesiaci užívania.

- a) Učinnosť pripravku preskumajte vhodnými obrázkami.
- b) Otestujte spomínanú famu o príberaní.
- c) Z rozdelenia s akým sklonom zrejme pochádzajú hmotnosti pred užívaním? Uvedte aspoň 2 dôvody.

20. **www.weather.com**

Internetová stránka `www.weather.com` sa snaží predpovedať počasie pre celú zemegulú. V súbore `weatherDOTcom.txt` je predpovedaná maximálna teplota pre Bratislavu (stĺpec `predpoved`) počas náhodne sledovaných 50 dní a v stĺpci `skutocnost` je maximálna hodnota, ktorú v ten deň v Bratislave nameral SHMU.

- a) Otestujte, či možno vo všeobecnosti tvrdiť, že `www.weather.com` predpovedá v priemere vyššiu teplotu, než je potom skutočnosť.
- b) Z rozdelenia s akým sklonom zrejme pochádzajú údaje v stĺpci `skutocnost`? Uvedte aspoň 2 dôvody.

21. **Vrh guľou**

Vrháč guľou si do súboru `vrh_gulou.txt` počas niekoľkých tréningových dní zapisoval dátum (stĺpec `datum`), maximálny dosiahnutý výkon v metroch (stĺpec `vykon`) a to, či v ten deň pršalo (stĺpec `dazd`).

- a) Samostatne pre daždové a samostatne pre bez daždové dni nájdite dátumy, kedy je možné výkon považovať za outlier. Z ďalších analýz tieto dni vylúčte.
- b) Otestujte túšenie, že dažď nepriaznivo pôsobí na vrháčov výkon. Výsledok ozrejmte „ľudskou recou“.

22. **Gorila**

Vydavateľstvo Petit Press si urobilo menší prieskum vplyvu písania o kauze Gorila na predajnosť ich denníka SME. V súbore `gorila.txt` sú v stĺpci `stanok` označené novinové stánky na niekoľkých autobusových staniciach na Slovensku: `december6` a `januar17` sú počty predaných SME v jeden bezný utorok pred a v jeden bezný utorok po vypuknutí Gorily.

- a) Nájdite stanok, v ktorom došlo k najväčšiemu percentuálnemu nárastu predajnosti SME.
- b) Otestujte, či je možné tvrdiť, že sa SME po prepuknutí Gorily predáva lepšie. Výsledok ozrejmte „ľudskou recou“.

23. Nedelne kazne

Nedelnu kazen v kostole ma niekedy pan farar a inokedy zase pan kaplan. Sedemrocneho ministranta ich retorika zatiaľ prilis neoslovuje, a tak si obcas krati čas meraním dĺžky kazni. V subore `kazne.txt` su uvedene dĺžky zopár „odstopovanych“ kazni (stĺpec `dĺzka`) a kto bol kazateľom (stĺpec `kto`).

- Samostatne pre kazdeho kazateľa najprv vylučte kazne, ktorých dĺžka je outlierom, a ďalšie analýzy robte bez týchto údajov.
- Otestujte pravdivosť legendy, ktorá sa koluje medzi miestnymi dochodkynami: vraj pan farar kazu dlhšie. Výsledok im ozrejmite „ľudskou recou“.

24. Rohliky

Zákazník chcel overiť, či Billa a Lidl predávajú klasické (údajne) 40-gramové rohlíky v priemere s rovnakou hmotnosťou. Raz za čas kúpil v náhodnom obchode jeden rohlík (vždy ten, čo bol v prvej nádobe a najviac na „juhozapade“), doma ho odvážil a do suboru `rohlíky.txt` o ňom zapísal príslušné údaje do stĺpcov `hmotnosť` (v gramoch), `obchod` a `datum`.

- Samostatne pre rohlíky z Billy a samostatne pre rohlíky z Lidlu nájdite dátumy, keď možno považovať zakúpený rohlík kvôli jeho hmotnosti za outliera. Z ďalších analýz tieto dni vylučte.
- Je možné tvrdiť, že Billa a Lidl sa priemernou hmotnosťou rohlíkov líšia? Výsledok ozrejmite „ľudskou recou“.

25. Prokrastinácia (Skutočné data)

Študenti EFM mali do 14. decembra 2014 do polnoci odoslať e-mailom domacu úlohu z Počítačovej štatistiky. V subore `prokrastinacia.txt` sa v stĺpci `čas` píše, koľko minút pred termínom študent úlohu poslal, pričom `NA`-čko znamená, že úlohu neposlal a záporné číslo znamená odoslanie po termíne. (Uhadnite, čo je v stĺpcoch `meno` a `pohlavie`.)

- Vypíšte mená ľudí, čo domacu úlohu vôbec neodovzdali a z ďalších výpočtov ich vynechajte.
- Zistite, kto odovzdal úlohu ako prvý a kto ako úplne posledný, no ešte nacas.
- Otestujte, či možno na základe získaných údajov tvrdiť, že matfyzácky su väčšie „prokrastinatorky“ než matfyzáci (čiže viac odkladajú povinnosti na neskoro), pričom do testovania zahrňte len tých, čo odovzdali úlohu nacas.

2 Pravdepodobnosť uspechu

26. Winnetou a Old Shatterhand

Winnetou a Old Shatterhand počas 30 dní strieľali na terč. V subore `winnetou_old_shatterhand.txt` je v stĺpcoch `strelyW` a `strelyOS` zaznamenané, koľkokrát v daný deň Winnetou resp. Old Shatterhand strieľali na terč. V stĺpcoch `zasahyW` a `zasahyOS` je zaznamenané, koľkokrát Winnetou resp. Old Shatterhand trafili terč.

Kto z nich sa zda byť lepší strelec? Svoje túsenie otestujte pomocou p-hodnoty i „thresholdov“.

27. Sebavedomy basketbalista

Basketbalista pred podpisom zmluvy v NBA o sebe vyhlasoval, že jeho úspešnosť pri trestných hodoch je 90%. Počas 30-dňového prípravného kempu hádzal každý deň na kos. V subore `basketbalista.txt` je v stĺpci `hody` uvedené, koľko trestných hodov v daný deň hádzal, a v stĺpci `zasahy` je uvedené, koľko z nich padlo do kosa.

Otestujte basketbalistov výrok pomocou p-hodnoty i „thresholdov“ a urobte 95%-ný interval spoľahlivosti pre skutočnú úspešnosť basketbalistu. „Ľudskou recou“ vysvetlite, čo tento interval hovorí.

28. Spolupráca s StB

Komunistická tajná služba StB urobila na východe tajný nábor spolupracovníkov. Vyslala svojich agentov do Kosíc a Presova, aby nahovarali ľudí na kolaboráciu. V subore `stb.txt` je v stĺpcoch `oslovenychK` resp. `oslovenychP` uvedené, koľko ľudí agenti v daný deň oslovili v Kosiciach resp. v Presove. V stĺpcoch `ziskanychK` resp. `ziskanychP` je uvedené, koľko z týchto oslovených ľudí aj získali na spoluprácu.

Ktore mesto sa zda byť náchylnejšie na spoluprácu z StB? Svoje túsenie otestujte a vysvetlite „ľudskou recou“.

29. Pes hľadá drogy

Polícia vycvčila psa na hľadanie drog v autách. V subore `pes_a_drogy.txt` sú uvedené údaje za 30 dní testovania. Stĺpec `auta` hovorí, koľko aut s drogami polícia v daný deň psovi „podstrečila“ a stĺpec `nasiel` hovorí, v koľkých z týchto aut pes drogu aj našiel. Kritéria EÚ hovoria, že na Schengenskú hranicu môže byť nasadený len pes s úspešnosťou aspoň 95%.

Ake túsenie máte z dát ohľadom možného nasadenia vášho psa? Otestujte to pomocou p-hodnoty i „thresholdov“. Výsledok skomentujte „ľudskou recou“. Urobte 95%-ný interval spoľahlivosti pre skutočnú úspešnosť psa.

30. Volvo a Skoda Favorit

Manžel jazdí na novom Volve a manželka na starej Skode Favorit. Vsimajú

si, ako často ich kontroluje polícia. V subore `volvo_vs_skoda.txt` je v stĺpcoch `volvo` resp. `skoda` uvedené, koľkokrát za posledných 20 týždňov stretlo policajnú hliadku Volvo a koľkokrát Skoda. V stĺpcoch `volvoZ` resp. `skodaZ` je uvedené, koľko z týchto stretnutí sa skončilo zastavením. Ako bývalí absolventi EFM si vypočítali, ktoré auto viac priťahuje pozornosť policajtov a svoje túženie otestovali. Čo im vyšlo? Výsledok skomentujte „ľudskou recou“.

31. Umely smiech

Televízia JOJ chce zistiť, či je potrebné vkladať umelý smiech (tzv. laugh track) do televízneho seriálu Mafstory. Najala si preto dobrovoľníkov, ktorým pustila 5 epizód tohto seriálu bez umeleho smiechu. V subore `smiech.txt` je v stĺpci `humor` uvedený počet humorných miest v danej epizóde a stĺpci `smiech` je uvedené, pri koľkých z týchto humorných miest sa dobrovoľníci zasmiali. Nasledne iným dobrovoľníkom pustili iných 5 epizód, ale tentoraz aj s umelým smiechom na humorných miestach (výsledky sú v stĺpcoch `humorPLUS` a `smiechPLUS`). Pomocou p-hodnoty i kritických hodnôt otestujte účinok umeleho smiechu a „ľudskou recou“ vysvetlite výsledok.

32. Maly Copperfield

Kuzelník David Copperfield sa už ako dieta na pieskovisku vychvaloval, že má telepatické schopnosti a je schopný predpovedať, či pri hode mincou padne líce alebo rub. Vždy sa našlo nejakej diete, ktoré bolo ochotné hadzať mincou. Výsledky z 15 dní sú v subore `copperfield.txt`. V stĺpci `hody` je počet hodov mincou v daný deň a v stĺpci `spravne` je uvedené, koľkokrát David uhadol výsledok. Môže David na základe uvedených historických dát tvrdiť, že má nadprirodzené schopnosti (teda že jeho úspešnosť je nad 50%, čo má v priemere každý)? Test vykonajte pomocou p-hodnoty i pomocou kritických hodnôt. Určte 95%-ný interval spoľahlivosti pre skutočnú Davidovu úspešnosť.

33. Lasak vs. Stana

Blížia sa majstrovstvá sveta v hokeji a tréner Jan Filc rieši brankársku otázku. Sleduje zápasy, v ktorých za svoje kluby chytajú „naši chlapci“ Jan Lasak a Rastislav Stana. Do suboru `lasak_stana.txt` si zapisuje, či v danom zápase chytal Lasak alebo Stana, koľko striel vyslal súper na branku a koľko dostal daný brankár gólov.

- Zistíte, koľko „shutoutov“ zaznamenal Stana a koľko Lasak (shutout = brankár nedostal v zápase gól).
- Otestujte tvrdenie Stanovho manažera v televízii, že jeho klient má úspešnosť zásahov 94% (úspešnosť zásahov = pomer chytených striel ku všetkým streľam na branku). Výsledok objasnite „ľudskou recou“.

- c) Który brankar miał w sledowanych zapasach celkovo vyššiu úspešnosť zasahov? Otestujte, či ho možno ho na základe toho považovať za lepšího, než je jeho konkurent. Výsledok objasnite „ľudskou recou“.

34. Tildenov gem

V tenise sa hrajú tzv. sety, ktoré sa skladajú z tzv. gemov. Siedmy gem setu sa nazýva Tildenov (podľa slávneho amerického tenistu Billa Tildena) a pripisuje sa mu priam magický význam: vraj kto ho vyhra, mal by už vyhrať celý set. V subore `Tildenov_gem.txt` sú údaje z posledných 10 ročníkov turnaja vo Wimbledone: v stĺpci `TildenM` je uvedené, koľkokrát sa hral Tildenov gem v mužskej časti turnaja, a v stĺpci `VyhryM` sa uvádza, koľkokrát víťaz Tildenovho gemu následne vyhral celý set (analogický význam majú `TildenZ` a `VyhryZ` pre ženskú časť turnaja). Komu viac psychicky pomáha výhra Tildenovho gemu: mužom, či ženám? Svoje túšenie otestujte a výsledok objasnite „ľudskou recou“.

35. Stavky v Nike

Stavková spoločnosť Nike zaznamenávala počas jedného mesiaca do suboru `nike.txt`, koľko tiketov sa podalo v daný deň v Bratislave resp. v Trnave (stĺpce `stavkyBA` a `stavkyTT`) a koľko z nich bolo nakoniec výherných (stĺpce `vyhryBA` a `vyhryTT`). V ktorom meste sa zdajú byť úspešnejší típeri? Svoje túšenie otestujte, pričom test vykonajte pomocou p-hodnoty i pomocou kritických hodnôt.

36. Futbalové remízy (Skutočné data. Idea: Tomas Molokac.)

Vo futbale sa kedysi udeľovali za víťazstvo 2 body (za remízu 1 a za prehru 0). Často sa však stávalo, že v prípade nerozhodného stavu sa mužstvá nesnažili dať gól a uspokojili sa s remízou. Z dôvodu zvýšenia atraktívnosti zápasov sa preto v rokoch 1994 a 1995 v mnohých futbalových ligách v Európe zaviedlo pravidlo 3 bodov za víťazstvo. Údaje z niekoľkých krajín sú v subore `futbalove_remizy.txt`. V stĺpci `sezona` je uvedené, či ide o futbalovú sezónu tesne pred alebo tesne po zavedení pravidla o 3 bodoch. Údaj `zapasy` znamená počet zápasov v danej sezóne a údaj `remiz` znamená počet remíz v danej sezóne. Je výskyt remíz po reforme väčší? Túšenie otestujte pomocou kritických hodnôt i pomocou p-value a výsledok vysvetlite „ľudskou recou“.

37. Erotické obrázky

Denník SME priniesol 11. januára 2011 správu, že americkí vedci vykonali experiment, počas ktorého postupne posadili pred 2 tmavé obrázky dobrovolníkov. Dobrovolník mal využiť svoje jasnovidecké schopnosti a stanoviť, na ktorej obrázku sa o chvíľu zjaví erotický obrázok. Následne pociťac náhodne vybral jednu z obrázkov a ten obrázok na nej ukázal. Ak sa dobrovolníkov

tip ukazal ako spravny, tak do suboru `eroticke_obrazky.txt` sa zapisalo „UHADOL“, inak „NEUHADOL“.

Dennik SME napisal, ze vysledky nie su z pohladu dokazania jasnovidec-tva presvedcive. Otestujte to, pricom test vykonajte pomocou p-hodnoty i pomocou kritickych hodnot. Urcte 95%-ny interval spolahlivosti pre ludsku schopnost predpovedat polohu erotickych obrazkov. Vysledky objasnite „ludskou recou“.

38. Trest smrti

Agentura pre vyskum verejnej mienky zistovala, aka je na Slovensku podpora trestu smrti. V subore `trest_smrti.txt` su zozbierane udaje od anketarov: stlpec `opytal` hovorí, kolko ludi danemu anketarovi odpovedalo, stlpec `ano` hovorí, kolko z nich s trestom smrti suhlasí, a v stlpci `kraj` je uvedene, ci sa dany anketar pytal v Nitrianskom alebo v Trnavskom kraji. Otestujte, ci je medzi uvedenymi krajmi rozdiel v podpore trestu smrti. Test vykonajte pomocou p-hodnot i kritickych hodnot. Skonstruujte obojstranny interval spolahlivosti pre rozdiel skutocnych preferencií pre trest smrti v tychto kra-joch a vysledok ozrejmite „ludskou recou“.

39. Alkohol za volantom

Polica jti v Bratislave a Stropkove robili pocas 20 sobotnajsich noci nahodne kontroly na alkohol u vodico v. V tabulke `alkohol_za_volantom.txt` su v stlpci `konroly` pocty kontrol za noc, v stlpci `alkohol` je udaj, kolko z tych kontrol zistilo v dychu vodica alkohol, a v stlpci `mesto` je napisane, ci ide o udaje z Bratislavy alebo zo Stropkova. Možno na zaklade tychto udajov vo vseobecnosti tvrdit, ze bratislavski vodici su menej zodpovedni nez stropkov-ski? Urcte aj obojstranny interval spolahlivosti pre rozdiel ich mier nezod-povednosti.

40. Klicenie semiacok

Skumala sa klicivost semiacok patizonu, pricom sa pouzilo sadivo od firmy Semena Veleliby alebo Semo Smrzice. V subore `klicenie_semiacok.txt` su udaje z 35 pokusnych policok: stlpce `zasadene` a `vyklicilo` hovoria, kolko sa na dane policko zasadilo semienok a kolko z nich aj vyrastlo. Stlpec `znacka` oznacuje vyrobcu sadiva. Ktore sadivo sa zda mat lepsiu klicivost? Otestujte to, pricom test vykonajte pomocou p-hodnoty i pomocou kritickych hod-not. Urcte 95%-ny interval spolahlivosti pre rozdiel skutocnych klicovosti. Vysledky objasnite „ludskou recou“.

41. Lietadlo „stealth“

Americka armada vraj vyvinula novy model lietadla s technologiou „stealth“, ktora znizuje sancu na jeho zachytenie radarom. Boli vykonane testy, pocas ktorych 15 dni prelietavali v istej pevnej vzdialenosti od radaru novy alebo

stary model lietadla „stealth“. Z testov unikol dokument `stealth.txt`, v ktorom stlpce `overflightsOLD` a `overflightsNEW` hovoria o počte preletov a stlpce `detectedOLD` a `detectedNEW` o príslušných počtoch radarových zachytení stareho resp. noveho typu lietadla v daný deň. Otestujte pomocou p-hodnoty i kritických hodnôt, či je možné tvrdiť, že nový typ lietadla je radarmi ľahšie zachytiteľný. Určte aj intervalový odhad pre rozdiel sancí zachytenia daným radarom na danú vzdialenosť. Výsledky objasnite „ľudskou recou“.

3 Korelacia

42. PlayStation 3 (I.)

Firma SONY uviedla na ázijský trh hernú konzolu PlayStation 3 a uvedenie podporila reklamnou kampanou. V subore `sony_azia.txt` sú údaje pre 45 ázijských štátov: v stĺpci `reklama` sú reklamné výdavky a v stĺpci `tržba` sú tržby za prvý mesiac predaja.

Otestujte túšenie, že skutočný korelačný koeficient medzi výdavkami na reklamu a tržbami je kladný a určte pre 95%-ný interval spoľahlivosti.

43. PlayStation 3 (II.)

Firma SONY uviedla na ázijský a európsky trh hernú konzolu PlayStation 3 a uvedenie podporila reklamnou kampanou. V subore `sony_azia_europa.txt` sú údaje pre niekoľko ázijských a európskych štátov: v stĺpci `reklama` sú reklamné výdavky, v stĺpci `tržba` sú tržby za prvý mesiac predaja a v stĺpci `svetadiel` je uvedené, z akého kontinentu je daná krajina.

Na ktorom kontinente sa zda byť skutočný korelačný koeficient medzi výdavkami na reklamu a tržbami väčší? Svoje túšenie otestujte a výsledok skomentujte „ľudskou recou“.

44. Diváci na futbale

V subore `navstevnost.txt` sú uvedené bodové zisky a priemerne počty divákov anglických prvotriednych klubov za sezónu 1992/93 a za sezónu 2006/2007. Pomocou korelačných koeficientov zistíte, či po 14 rokoch je silnejší/slabší súvis medzi kvalitou hry a navštevnosťou. Svoje túšenie otestujte a výsledok skomentujte „ľudskou recou“.

45. eRko vs. nelinearko

57 študentov EFM sa zúčastnilo na skúškach z eRka a nelinearka. V subore `erko_vs_nelinearko.txt` sú percentuálne zisky jednotlivých študentov.

- a) Pomocou hrádieb zistíte, koľko je outlierov medzi percentami z eRka a koľko je outlierov medzi percentami z nelinearka. V úlohách b)-d)

neberte do uvahy studentov, ktorí su outliermi v aspon jednom z predmetov.

- b) Otestujte tusenie, ze medzi vysledkami z eRka a vysledkami z nelinearka je pozitivny suvis.
- c) Urcte interval spolahlivosti pre skutocny korelacny koeficient medzi vysledkami z eRka a vysledkami z nelinearka.
- d) Otestujte tusenie, ze skuska z eRka je lahsia.

46. NHL

V subore `nhl.txt` su uvedene vydavky na platy hracov a ziskane body klubov hokejovej sutaze NHL v sezone 2006/2007. Zistite, ci je mozne tvrdit, ze medzi platmi a bodmi je pozitivny suvis a vypocitajte 95%-ny interval spolahlivosti pre skutocny korelacny koeficient medzi tymito velicinami.

47. NHL: vychod vs. zapad

V subore `nhl_vychod_zapad.txt` su pre kluby hokejovej sutaze NHL uvedene za sezonu 2006/2007 vydavky na platy hracov, ziskane body a prislusnost do konferencie. (Co je to konferencia: kluby NHL su pocas sezony rozdelené do 2 skupin, ktore sa nazývajú konferencie.) V ktorej konferencii sa zda byt skutocny korelacny koeficient medzi prijmami a ziskanyimi bodmi vacsi? Svoje tusenie otestujte a vysledok ozrejmite „ludskou recou“.

48. Prospech a plat

U absolventov EFM a FEI STU, ktorí isli po studiu do praxe, sa skumal suvis medzi priemernym prospechom absolventa pocas studia a jeho cistym nastupnym platom. Uda je su tabulke `prospech_a_plat.txt`, pricom v udaj stlpci `prospech` sa ziskal prevodom pismenok: A=1; B=1,5;...; E=3.

- a) Na absolventoch z EFM otestujte vyrok: „Prospech mi bude aj tak k nicomu.“
- b) Pomocou korelacnych koeficientov otestujte tusenie, ze u EFM prospech viacej vplyva na nastupny plat nez u FEI.

49. Kupaliska

Cez letne prazdniny sa pocas bezdazdovych dni sledovala maximalna teplota vzduchu a navstevnost na nekrytych kupaliskach v Diakovciach a v Londyne. V subore `kupaliska.txt` su prislusne udaje v stlpcoch `teplota`, `ludia` a `kupalisko`. Jeden riadok znamena jeden bezdazdovy den. Pomocou korelacnych koeficientov otestujte, ci mozno tvrdit, ze Anglicania su pri rozhodovaní sa o navsteve kupaliska menej ovplyvnovani teplotou vzduchu nez Slovaci.

50. Studený maj, v stodole raj

Podľa známej pranostiky by malo chladne (a tým padom daždive) počasie v máji blahodarné posobiť na úrodu. V subore `studený_maj.txt` sú v stĺcoch `teplota` a `uroda` uvedené priemerné majové teploty a priemerné výnosy obilí (v tonách na hektár) za posledných 40 rokov. Ako je uvedené v stĺpci `družstvo`, údaje pochádzajú z poľnohospodárskych družstiev v Marcelovej na juhu Slovenska (nepárne roky) a v Trstenej na severe Slovenska (párne roky).

- Platnosť pranostiky overte na dátach z Trstenej.
- Je možné tvrdiť, že na juhu Slovenska neplatí pranostika rovnako silno ako na severe? Test vykonajte pomocou p-hodnoty i kritických hodnôt.

BONUS: Preto nesmú údaje z Marcelovej a Trstenej pochádzať z rovnakých rokov?

51. Nezamestnanosť vs. voľby (Skutočné data.)

V subore `nezamestnanosť_volby.txt` sú údaje o percente nezamestnaných (stĺpec `nezamestnanosť`) v 79 okresoch Slovenska v roku 2010. Stĺpec `podpora` udáva pre každý okres percentuálny zisk strany ****-** v parlamentných voľbách konaných v tom istom roku.

- Na údajoch z celého Slovenska otestujte túšenie, že podpora danej strany pozitívne súvisí s mierou nezamestnanosti.
- Je možné tvrdiť, že v Bratislavskom kraji je súvis medzi nezamestnanosťou a podporou danej strany silnejší než v Kosickom?

BONUS: Čo ukryva ** za pomlčkou v názve strany? DS či SD? :-)

52. Dyktat

Vedci zistovali vplyv alkoholu na počet gramatických chýb pri dyktate. Skusobným diktovaním v strave triedením našli 50 dobrovolníkov z približne rovnakou náchylnosťou k chýb/prklep;?!. pOTOM každého opýyli na yní stupeň opilosti (subor `dyktat.txt`, stĺpec `promile`) a nadiktovali ym tetx v korrom každemu sral počet chýb (spec `chyby`). Stĺpce `pohlavie` hovorí ic to je žena alebo no žena;

- oestujte na údaje od ziiieen či je ozaj suuuvys meci počtom chýb a prommle pijatiky v zylach.
- mozeme tvrdiť že u zein je vo vsouobecnosti sylneejs súvys medzi počtom chýb a opitocstouuu než u mužov? aj trizvou „luckov recou“ vysvetlite... koniec halasenia bodka

4 Linearna regresia

53. Cyklisti a EPO

Cyklistický klub T-Mobile mal 20 cyklistov a ti mali zhruba rovnaku vykonnosť, ktorá sa meria „hodinovkou“ (=koľko kilometrov cyklista prejde za hodinu). Potom začali užívať doping EPO. V súbore `cyklisti_epo.txt` je zaznamenané, koľko dní dany cyklista užíval EPO (=stĺpec `dni`) a koľko kilometrov potom zvladol prejsť v hodinovke (=stĺpec `kilometre`).

- Zostavte lineárny model závislosti výkonnosti cyklistu od doby užívania látky EPO. „Ludskou recou“ vysvetlite, čo znamenajú parametre modelu.
- Bez použitia funkcie `summary` otestujte hypotézu, že užívanie látky EPO nemá vplyv na výkonnosť.
- Určte 95%-ný interval pre strednú hodnotu počtu prejdenej kilometrov v hodinovke, ak sa EPO užíva 30 dní. „Ludskou recou“ vysvetlite, čo tento interval znamená.
- Vieme, že cyklista Jan Ullrich užíval EPO najmenej 30 a najviac 40 dní. Práve sa chystá absolvovať hodinovku. Určte interval, v ktorom možno očakávať počet kilometrov, ktoré v hodinovke prejde. „Ludskou recou“ vysvetlite, čo tento interval znamená.

54. Zber jahod

Jahodová farma v Anglicku využívala toto leto na zber jahod výhradne poľských a slovenských brigádnikov. Každé ráno vyhnal majiteľ na pole zopár Poliakov a Slovakov a večer odvažil, koľko kilogramov spolu nazbierali. V súbore `zber_jahod.txt` je v stĺpcoch `poliaci` a `slovaci` zapísané, koľko Slovakov a Poliakov v daný deň pracovalo, a v stĺpci `jahody` je uvedené, koľko kilogramov jahod v ten deň nazbierali.

- Zostavte lineárny model popisujúci situáciu a „ludskou recou“ vysvetlite, čo hovoria odhadnuté parametre.
- Otestujte, či je v modeli splnený predpoklad normality.
- Kto sa podľa dát zda byť usilovnejší: Poliaci či Slovaci? Svoje tušenie otestujte.

55. Tazba v bani

Je známe, že množstvo vytazenej horniny v bani lineárne závisí od tvrdosti horniny, teploty v bani, počtu pracujúcich baníkov a vlhkosti vzduchu. V súbore `bana.txt` sú uvedené údaje za 50 dní ťažby.

- a) Zostavte linearny model popisujuci situaciu a „ludskou recou“ vysvetlite, co hovoria odhadnute parametre.
- b) Bez pouzitia funkcie `summary` vypocitajte koeficient determinacie modelu a „ludskou recou“ vysvetlite, co toto cislo hovori.
- c) Otestujte hypotezu, ze mnozstvo vytazenej horniny nezavisí od teploty ani vlhkosti v bani.

56. SuperStar

Hudobne vydavatelstvo SONY BMG vydalo 40 albumov finalistov SuperStar v Cechach a na Slovensku. Kazdy album podporilo reklamnou kampanou. V subore `superstar.txt` je v stlpci `reklama` zapisane, kolko tisíc korun sa minulo pri danom albume na reklamu, v stlpci `predane` je zapisane, kolko kusov CD sa predalo, a v stlpci `pohlavie` je uvedene, ci album naspievala zena (=“Z”) alebo muz (=“M”). Je známe, ze predajnost albumov linearne zavisí od reklamnych vydavkov.

- a) Zostavte linearny model popisujuci tuto zavislost pri zenskyh albumoch a linearny model popisujuci tuto zavislost pri muzskyh albumoch. „Ludskou recou“ vysvetlite vyznam parametrov modelov.
- b) Do obrazku zakreslite udaje pre muzske a zenske albumy a nakreslite odhadnute regresne priamky.
- c) V ktorom modeli sa zda byt skutocna regresna priamka strmsia? Tusenie otestujte a vysledok skomentujte „ludskou recou“.

57. Skuska z eRka

54 studentov EFM sa pripravovalo na skusku z eRka. V subore `skuska_z_eRka.txt` je uvedene, kolko hodin sa ucili v noci, kolko hodin cez den a kolko nakoniec ziskali na pisomke bodov.

- a) Zostavte linearny model popisujuci vzťah medzi ucením a pocetom bodov a „ludskou recou“ vysvetlite, co hovoria odhadnute parametre.
- b) Bez pouzitia funkcie `summary` vypocitajte koeficient determinacie modelu a „ludskou recou“ vysvetlite, co toto cislo hovori.
- c) Je efektivnejsie ucit sa v noci alebo cez den? Svoje tusenie otestujte.
- d) Student sa ucil 6 hodin v noci a 2 hodiny cez den. Prave kraca nevyspaty popri cintorinu na skusku. V akom intervale moze ocakavat svoj bodovy zisk? „Ludskou recou“ vysvetlite, co tento interval znamena.

58. Prasok ARIEL

Firma Procter&Gamble uviedla na trh nový druh prasku ARIEL a počas prvého týždňa ho podporila reklamnou kampanou v 50 európskych krajinách. V subore `ariel.txt` je v stĺpci `billboardy` uvedené, koľko billboardov prenajali, v stĺpci `tlac` je počet reklamných cm^2 v novinách a časopisoch a v stĺpcoch `televizia` resp. `radio` je počet zakúpených reklamných minút v televíziách resp. v rádiách. Stĺpec `trzb` uvádza tržbu za prvý týždeň (v tisícoch) v danej krajine prepocítanu na milión obyvateľov.

- Zostavte lineárny model popisujúci predajnosť a „ľudskou recou“ vysvetlite, čo hovoria odhadnuté hodnoty parametrov.
- Otestujte hypotézu, že reklama v tlači a v rádiách nemá vplyv na predajnosť.
- Určte interval spoľahlivosti pre strednú hodnotu tržby, ak Procter&Gamble na prvý týždeň objedná 60 billboardov, 2000 cm^2 reklamy v tlači, 200 minút reklamy v televíziách a 20 minút reklamy v rádiách.
- Otestujte, či je v modeli splnený predpoklad normality. Ak by nebol splnený: aké by to malo následky?

BONUS: Týždenný prenájom jedného billboardu stojí 10 100 korún. Otestujte, či sa Procter&Gamble oplatí propagovať ARIEL na billboardoch, ak je v krajine milión obyvateľov.

59. Uroda obilia

Výskumný ústav poľnohospodársky vysadil obilie na 40 lanoch a každý lan následne inak zavlažoval. V subore `obilie.txt` je v stĺpci `polievanie` uvedené, koľko hektolitrov vody sa počas vegetačného obdobia vylialo na 1 m^2 daného lanu. V stĺpci `uroda` je uvedené, koľko ton obilia sa urodilo (v prepote na 1 hektar) na danom lane.

- Zostavte lineárny model popisujúci závislosť medzi zavlažovaním a urodou. Nakreslite zodpovedajúci obrazok a odhadnutú regresnú priamku. „Ľudskou recou“ vysvetlite, čo hovoria odhadnuté hodnoty parametrov.
- Bez použitia funkcie `summary` vypočítajte koeficient determinácie modelu a „ľudskou recou“ vysvetlite, čo toto číslo hovorí.
- Otestujte hypotézu, že 1 hektoliter vody na m^2 zvýši úrodu v priemere o 0,5 tony obilia na hektar.

- d) Vieme, že na jeden z lanov sa počas vegetačného obdobia vylialo aspoň 0,4 a najviac 0,8 hektolitrov vody na m^2 . Určte interval, v ktorom možno na 95% očakávať úrodu obilí.

60. Spotreba Peugeotu (idea: Martin Svorad)

Rodina Divných si kúpila auto Peugeot 207. Keďže ho vyrobili Slováci, rodina nedôveruje výrobcovým údajom o objeme nádrže a spotrebe, preto sa ich pokúsili odhadnúť: vždy natankovali nádrž doplna a po určitom počte kilometrov zmerali, koľko v nádrži ostalo benzínu. Údaje sú v súbore `peugeot.txt`. Stĺpec `kilometre` hovorí, koľko kilometrov auto od natankovania preslo, a stĺpec `nadrz` hovorí, koľko litrov benzínu zostalo v nádrži.

- a) Zostavte lineárny model popisujúci závislosť zostatku v nádrži od počtu prejdenej kilometrov. Nakreslite zodpovedajúci obrazok a odhadnutú regresnú priamku. „Ludskou recou“ vysvetlite, čo hovoria odhadnuté hodnoty parametrov.
- b) Otestujte tvrdenie výrobcu, že spotreba Peugeotu je 6,5 litra na 100 km.
- c) Matka cestuje Peugeotom do práce 21 km. Odhadnite strednú hodnotu výdavkov na benzín na 1 cestu tam a späť, ak liter benzínu stojí 0,93€. Určte aj 95%-ný interval spoľahlivosti.
- d) Rodina sa s autom ide pochváliť k starej mame. Vzdialenosť k nej nevedia presne, ale sú si istí, že je to medzi 150 a 170 km. Určte interval, v ktorom môžu na 95% očakávať objem zostávajúceho benzínu po dojazde (pred cestou bola nádrž plná).

61. Spotreba Peugeotu a Skody

Motoristická redakcia denníka SME overovala spotrebu Peugeotu 207 a Skody Fabia. Počas 30 dní urobili každý deň 1 jazdu na Peugeotu i Skode. Pred jazdou bola v autách vždy plná nádrž. Po jazde zapísali do súboru `peugeot_skoda.txt` počet prejdenej kilometrov (stĺpce `kilometreP` a `kilometreS`) a množstvo benzínu, čo ostalo v nádrži (stĺpce `nadrzP` a `nadrzS`).

- a) Zostavte lineárny model popisujúci závislosť množstva benzínu v nádrži od počtu prejdenej kilometrov u Peugeotu a lineárny model popisujúci túto závislosť u Skody. „Ludskou recou“ vysvetlite význam parametrov modelov.
- b) Do obrazku zakreslite údaje pre Peugeotu i Skodu a nakreslite odhadnuté regresné priamky.

- c) Vyrobcovia uvadzaju, ze Peugeot 207 a Skoda Fabia maju rovnaku spotrebu. Tvrdenie otestujte a vysledok skomentujte „ludskou recou“.

62. Mlade ninja korytnacky

Michelangelo, Raphael, Donatello a Leonardo maju vo zvyku raz za cas preplesknut zloducha Shreddera. V subore `korytnacky.txt` su udaje z 50 bitiek a v kazdom riadku je uvedene, kolko uderov ktora korytnacka Shredderovi ustrčila pri danej bitke. Okrem toho mu vzdy jednu ranu Donatellovou palicou uderil aj Majster Splinter. V stlpci PN je uvedene, kolko dni bol potom Shredder praceneschopny.

- a) Zostavte linearny model popisujuci vplyv uderov jednotlivych korytnaciek na praceneschopnost. „Ludskou recou“ vysvetlite, co hovoria odhadnute hodnoty parametrov.
- b) Majster Splinter machruje, ze i ked je stary, jeho udery su stale rovnako nicive ako Donatellove. Otestujte to.
- c) Otestujte hypotezu, ze Michelangelove a Raphaelove udery nesposobuju Shredderovi nijaku ujmu na zdravi.

63. Gumkaci

Medvedici Gumkaci piju gumidzus, ktory im umoznuje robit dlhe skoky. V tabulke `gumkaci.txt` je v stlpci `gumidzus` uvedene, kolko mililitrov gumidzusu medvedik vypil, v stlpci `vek` je uvedene, kolko ma medvedik rokov a v stlpci `dlzka` je uvedene, ake dlhy skok mu namerali.

- a) Zostavte linearny model popisujuci zavislost dlzky skokov od veku a mnozstva vypiteho gumidzusu. „Ludskou recou“ vysvetlite, co hovoria odhadnute hodnoty parametrov v modeli.
- b) Bez pouzitia funkcie `summary` vypocitajte koeficient determinacie a „ludskou recou“ vysvetlite, co nam jeho hodnota hovorí.
- c) Patrocny medvedik vypil stvrt litra gumidzusu. Urcte interval spolahlivosti pre strednu hodnotu dlzky jeho skokov.
- d) Otestujte, ci je v modeli splneny predpoklad normality. Ak by nebol splneny: ake by to malo nasledky?

64. Kiribati

Ostrovny stat Kiribati, ktoremu hrozi zalatie morom, ma od zaciatku roka 1982 nove zariadenie na meranie vysky hladiny. V subore `kiribati.txt` je v stlpcoch `vyska` a `rok` uvedene, aky bol priemer nameranych denných vysok (v centimetroch) v roku $1982 + \text{rok}$. Ukazuje sa, ze vyska hladiny mora linearne narasta s casom.

- a) Zostavte linearny model popisujuci tuto zavislost a „ludskou recou“ vysvetlite vyznam jeho parametrov.
- b) Vhodnym obrazkom i testom overte, ci je v modeli splneny predpoklad normality. Co by sa stalo, ak by nebol?
- c) Mozno nasimi datami popriet tvrdenie, ze hladina oceanov sa zvysuje rocne o 2 cm? Odpoved podlozte testom.
- d) Odhadnite, aku hodnotu si Kiribatia zapisu do tabulky v roku 2020 a vypocitajte aj intervalovy odhad.

65. Prasata

Dve prasata boli od 10. tyzdna zivota krmene krmivami DYNAPHOS I resp. DYNAPHOS II. V subore `prasata.txt` je v stlpcoch `hmotnost` a `tyzden` uvedene, kolko vazilo dane prasa po danom pocte tyzdnov krmenia. V stlpci `krmivo` je uvedene, akym krmivom sa prasa krmilo. Predpokladame, ze hmotnost prasata linearne zavisí od doby krmenia.

- a) Zostavte linearny model popisujuci tuto zavislost pri krmive DYNAPHOS I a linearny model popisujuci tuto zavislost pri krmive DYNAPHOS II. „Ludskou recou“ vysvetlite vyznam parametrov modelov.
- b) Do obrazka zakreslite udaje pre prve i druhe prasa a nakreslite odhadnute regresne priamky.
- c) Ktore krmivo sa zda byt kvalitnejšie? Svoje tusenie otestujte a vysledok objasnite „ludskou recou“.

66. Bicyklom do prace: Epizoda I

Anton chodi kazdy den bicyklom do prace 9 km. V tabulke `bicyklom_do_prace_I.txt` su zapisane udaje z 50 nahodne vybranych jazd. Stlpec `cas` hovorí, kolko minut mu trvala cesta, stlpec `vietor` hovorí, aky silny protivietor fukal (zaporne hodnoty znamenaju vietor do chrbta), stlpec `teplota` hovorí, aka bola rano teplota a stlpec `pauza` znamena, kolko minut po ranajkach nasadol na bicykel. Predpokladame, ze doba jazdy linearne zavisí od spominanych premennych.

- a) Zostavte linearny model popisujuci tuto zavislost a „ludskou recou“ vysvetlite hodnoty odhadnutych parametrov.
- b) Bez pouzitia funkcie `summary` vypocitajte koeficient determinacie modelu a „ludskou recou“ vysvetlite, co toto cislo hovorí.

- c) Otestujte Antonovo tuesenie, ze teplota a dlzka pauzy po ranajkach nemaju vplyv na dosiahnuty cas.

67. Bicyklom do prace: Epizoda II

Anton chodi kazdy den bicyklom do prace 9 km. V tabulke `bicyklom_do_prace_II.txt` su zapisane udaje z 50 nahodne vybranych jazd. Stlpec `cas` hovorí, kolko minut mu trvala cesta, stlpec `vietor` hovorí, aky silny protivietor fukal (zaporne hodnoty znamenaju vietor do chrbta), stlpec `teplota` hovorí, aka bola rano teplota a stlpec `pauza` znamena, kolko minut po ranajkach nasadol na bicykel. Predpokladame, ze doba jazdy linearne zavisí od spominanych premennych.

- a) Zostavte linearny model popisujuci tuto zavislost a „ludskou recou“ vysvetlite hodnoty odhadnutych parametrov.
- b) Ktory riadok udajov v tabulke najviac ovplyvnuje odhady parametrov? Odhadnite parametre bez pouzitia tohto riadku a vyhodnotte situáciu. Podla vlastneho zvazenia pouzite v dalsich ulohach upravene odhady parametrov.
- c) Teplomer ukazuje -10°C a Jozef Ilko v radiu prave predpoveda severo-zapadny vietor minimalne 2 a maximalne 5 metrov za sekundu, co pre Antona znamena protivietor. 15 minut po ranajkach sada na bicykel. V akom intervale moze ocakavat cas v to rano? (Predpokladame, ze sila vetra sa pocas jazdy nemení.)
- d) Vhodnym obrazkom i testom overte, ci je v modeli splneny predpoklad normality. Co by sa stalo, ak by nebol?

68. Bicyklom do prace: Epizoda III

Anton chodi kazdy den bicyklom do prace 9 km. V tabulke `bicyklom_do_prace_III.txt` su zapisane udaje z 50 nahodne vybranych jazd. Stlpec `cas` hovorí, kolko minut mu trvala cesta, stlpec `vietor` hovorí, aky silny protivietor fukal (zaporne hodnoty znamenaju vietor do chrbta), stlpec `teplota` hovorí, aka bola rano teplota a stlpec `pauza` znamena, kolko minut po ranajkach nasadol na bicykel. Predpokladame, ze doba jazdy linearne zavisí od spominanych premennych.

- a) Zostavte linearny model popisujuci tuto zavislost a „ludskou recou“ vysvetlite hodnoty odhadnutych parametrov.
- b) Anton si povedal, ze vzdy, ked rano nebude fukat vietor a bude 15°C , tak okamzite po ranajkach vystartuje a zapise si dosiahnuty cas. K akemu cislu sa bude dlhodobo blizit priemer tychto zapisanych casov?

Určte aj intervalový odhad pre túto hodnotu a výsledok ozrejmte „ľudskou recou“.

- c) Otestujte, či je pre Antona horsie zosilnenie protivetra o 3 metre za sekundu alebo pokles teploty o 4°C .

69. A je to

Pat a Mat zo serialu „A je to“ sa začali venovať opravám meskajúcich hodín. S suborom `a_je_to.txt` sú údaje o 35 hodinách, čo po ich zásahu este fungovali: stĺpce `Pat` a `Mat` uvádzajú, koľko minút sa ktorý v hodinách vrtal. Stĺpec `zvyseniemeskania` hovorí, o koľko minút za deň sa po oprave zvýšilo meskanie hodín. Ukazuje sa, že toto číslo lineárne závisí od oboch časov opravovania a navyše pred zásahom samotných majstrov hodinky este minútu nemotorne otvára Maxipes Fík.

- a) Zostavte lineárny model popisujúci túto závislosť a „ľudskou recou“ vysvetlite význam jeho parametrov.
- b) Overte túšenie, že za rovnaku chvíľku hodinám uskodí menej Pat než Mat.
- c) Na stole ležia hodinky, ktoré meskajú 2 minúty za deň. Pat a Mat sa dohodli, že každý na nich pomajstruje 15 minút. Určte interval, v ktorom možno očakávať meskanie hodín po oprave.
- d) Vhodným obrazkom i testom overte, či je v modeli splnený predpoklad normality. Čo by sa stalo, ak by nebol?

70. Serif a vraždy

V subore `americke_staty_mini.txt` je zopár údajov zo scítania ľudu v USA v 70. rokoch. V stĺpci `Murder` je počet vražd za rok na 100 000 obyvateľov, v stĺpci `Income` je priemerný ročný príjem, v stĺpci `Illiteracy` je percentuálny podiel negramotných a v stĺpci `Frost` je priemerný počet mrazivých dní za rok.

- a) Zostavte lineárny model popisujúci závislosť počtu vražd od ostatných premenných a „ľudskou recou“ vysvetlite hodnoty odhadnutých parametrov.
- b) Bez použitia funkcie `summary` vypočítajte koeficient determinácie modelu a „ľudskou recou“ vysvetlite, čo toto číslo hovorí.
- c) V mestečku s 20 000 obyvateľmi sa serif ujal svojho uradu. Vie, že v danej oblasti nikdy nemrzne, priemerne ročné príjmy sú 3 700\$ a negramotnosť je 2,2%. V akom intervale môže očakávať počet vražd v mestečku za budúci rok?

- d) Otestujte tusenie, ze pocet mrazivych dni a prijmy su v spominanom modeli zbytocne.

71. Bufety pred stadionmi

Pred stadionmi Slovana Bratislava na bratislavských Pasienkoch a pred stadionom Spartaka v Trnave su dva sortimentom vcelku podobne bufety. V subore `bufety_pred_stadionmi.txt` su udaje z dni, ked sa na danom stadione hral domaci zapas uplynuleho rocnika ***** ligy: stlpec `divaci` uvadza pocet divakov, stlpec `trzba` je trzba bufetu v dany den a stlpec `stadion` hovorí, o ktory bufet sa jedna. Prevadzkovatelia bufetov vytusili, ze trzba linearne zavisi od poctu divakov.

- a) Pre kazdy z bufetov zostavte linearny model popisujuci tuto zavislost a „ludskou recou“ vysvetlite vyznam odhadnutych parametrov.
- b) Do jedneho obrazka nakreslite udaje pre oba bufety a odhadnute regresne priamky.
- c) V oboch modeloch overte vhodnym obrazkom a testom splnenie predpokladov normality. Ake nasledky by malo ich nesplnenie?
- d) V ktorom meste mina priemerny divak na obcerstvenie viacej? Svoje tusenie otestujte a vysledok ozrejmite „ludskou recou“.

72. Taxikar & majstrovstva sveta

Isty bratislavsky taxikar si zaumienil pocas celeho hokejoveho sampionatu vozit jedine Rusov a Finov. V subore `taxikar.txt` je pre kazdy z prvych 16 dni v stlpcoch `Fini` a `Rusi` zapisane, kolko vykonal „finskych“ a kolko „ruskych“ jazd. Stlpec `prepitne` uvadza, kolko prepitneho napocital pri vecernom uctovani.

- a) Zostavte linearny model popisujuci zavislost prepitneho od poctov „finskych“ a „ruskych“ jazd. „Ludskou recou“ vysvetlite vyznam jeho parametrov.
- b) Ktoru zo spominanych dvoch narodnosti sa viacej oplati rozvazat? Svoje tusenie otestujte.
- c) Vo finalovy den majstrovstiev taxikar cely den kmital ako fretka: 7-krat viezol Finov a az 8-krat Rusov. Prave sa cely zmorený vezie domov, no naladu mu zlepšuje pomyslenie na bliziacu sa zaverene pocitanie. V akom intervale moze ocakavat prepitne z tohto dna?
- d) Vhodnym obrazkom i testom overte, ci je v modeli splneny predpoklad normality. Co by sa stalo, ak by nebol?

73. Zbierka na charitu

Skupinky dobrovolníkov zbierali celý den pre po meste finančné príspevky pre charitatívnu organizáciu. Pred odchodom do terenu dostala každá skupinka do pokladničky pevnú (no pre nich neznámu) sumu, aby boli aj prví darcovia motivovaní. Dobrovolníci nepoznali presné hodnoty vhadzovaných príspevkov, no vždy si zaznamenali, či dal príspevok muž alebo žena. V súbore `zbierka_na_charitu.txt` sú v každom riadku údaje o jednej skupinke: v stĺpcoch `muži` resp. `ženy` sú počty mužských a ženských darcov a v stĺpci `vyzbierané` je uvedená celková vyzbieraná suma.

- Zostavte lineárny model popisujúci závislosť vyzbieranej sumy od počtov mužských a ženských darcov. „Ludskou recou“ vysvetlite význam jeho parametrov.
- Otestujte, či možno tvrdiť, že ženská prispievateľka je vo všeobecnosti strednejšia než prispievateľ-muž.
- Po scítaní darov sa ešte dovalila jedna oneskorená skupinka. Ich darcami bolo 72 mužov a 108 žien. Horúkovito sa pustajú do počítania. V akom intervale môžu očakávať výslednú sumu?
- Vhodným obrázkom i testom overte, či je v modeli splnený predpoklad normality. Čo by sa stalo, ak by niebol?

74. Sneženie

O piatej ráno začalo v Ziline i v Bratislave snežiť, pričom v tej chvíli bol v oboch mestách už i starý sneh. V každom meste snežilo celý den zhruba rovnomerne. Miestni amatérski meteorológovia však vyšli z postele a začali s nepravidelnými meraniami celkovej výšky snehu (starého+nového) až po ôsmej hodine ráno. V súbore `snezenie.txt` je pamiatka po tejto ich celodennej zábavke: stĺpec `čas` hovorí, koľko hodín po piatej ráno sa vykonalo dané meranie výšky snehu (stĺpec `výška`).

- Zostavte 2 lineárne modely: pre závislosť výšky snehu od času v Ziline a pre rovnakú závislosť v Bratislave. „Ludskou recou“ vysvetlite význam odhadnutých parametrov a situáciu zakreslite do jedného obrázka.
- Otestujte, či sú merania v súlade s tvrdením, že v Bratislave bolo pred snežením o 10 cm snehu menej než v Ziline.
- V Ziline snežilo až do neskorej večery, no tamojší meteorológ už bol príliš skrehnutý, aby vybiehal po ďalšie data. V akom intervale by sa dal očakávať výsledok, ak by zmeral výšku aj niekedy poobede medzi štvrtou a piatou hodinou?

75. Hod ostepom & doping

Vrhac ostepom si do suboru `hod_ostepom.txt` po pretekoch vzdy zapise rychlost vetra v metroch za sekundu (stlpec `vietor`; zaporna hodnota znamena protivietor), dlzku najdlhsieho hodu v metroch (stlpec `vykon`) a kolko dni predtym uzival 2 rozne druhy dopingu (stlpce `calusterone` a `androstenediol`).

- a) Zostavte linearny model popisujuci zavislost vykonu od vetra a dob uzivania dopingov a „ludskou recou“ vysvetlite vyznam jeho parametrov.
- b) Ktory riadok udajov v tabulke najviac ovplyvnuje odhady parametrov? Odhadnite parametre bez pouzitia tohto riadku a vyhodnotte situaciu. Podla vlastneho zvazenia pouzite v dalsich ulohach nove odhady parametrov.
- c) Otestujte, ci mozno tvrdit, ze dopovanie uvedenymi preparatmi nema vplyv na osteparov vykon.
- d) Otestujte, ci calusterone za rovnaku dobu uzivania ovplyvni vykon nazaj menej nez androstenediol.